Los Alamos HIV Immunology Database Overview

Bette Korber

**Summer School on Quantitative Systems Immunology,
Boston University
June 10-15, 2013**

HIV Immunology database PI: Karina Yusim

slides available as PDF documents:
http://www.hiv.lanl.gov/content/sequence/HIV/HIVWORKSHOP/index.html

Los Alamos HIV Immunology Database Overview

Bette Korber

Summer School on Quantitative Systems Biology,

Boston University

June 10-15, 2013

HIV Immunology database PI: Karina Yusim

Los Alamos
NATIONAL LABORATORY

Overview

Immunology database introduction
Epitope maps and epitope summary tables
T-cell epitope search
T-cell epitope variants
Antibody search
Features of most broadly neutralizing antibodies (antibody "A-list")
Neutralizing antibody resources: contexts and features – new
QuickAlign – Align an epitope to the HIV database alignments
N-glycosite – finds N-linked glycosylation sites
ELF – epitope location finder
Peptgen – list peptides for reagent development
HeatMap

HIV genome browser tools – coming soon

Mosaic Vaccine Maker, Epicover, and Posicover
*   - generate candidate vaccines*
*   - estimate epitope coverage*
*   - determine regional epitope coverage*

HIV Database workshop

Los Alamos
NATIONAL LABORATORY

Immunology Database Content

- Incorporates published HIV T cell (CTL, T-helper) epitope and Antibody information (emphasis on monoclonals)

- Key information regarding what is learned about epitopes and mAbs in each paper is included

- Types of data recorded:
  - ☐ Epitope sequence and location: HXB2 numbering, subtype
  - ☐ Natural infection or vaccine
  - ☐ Host HLA or MHC
  - ☐ Ab isotype, binding region, species
  - ☐ Notes summarize main findings

Los Alamos
NATIONAL LABORATORY

Immunology Database Statistics

- **Contents: data from 1985 through 2013**
  - 7441 CTL entries
  - 1315 T-helper entries
  - 2386 Ab entries
  - 3090 published citations

HIV Immunology Cumulative Entries
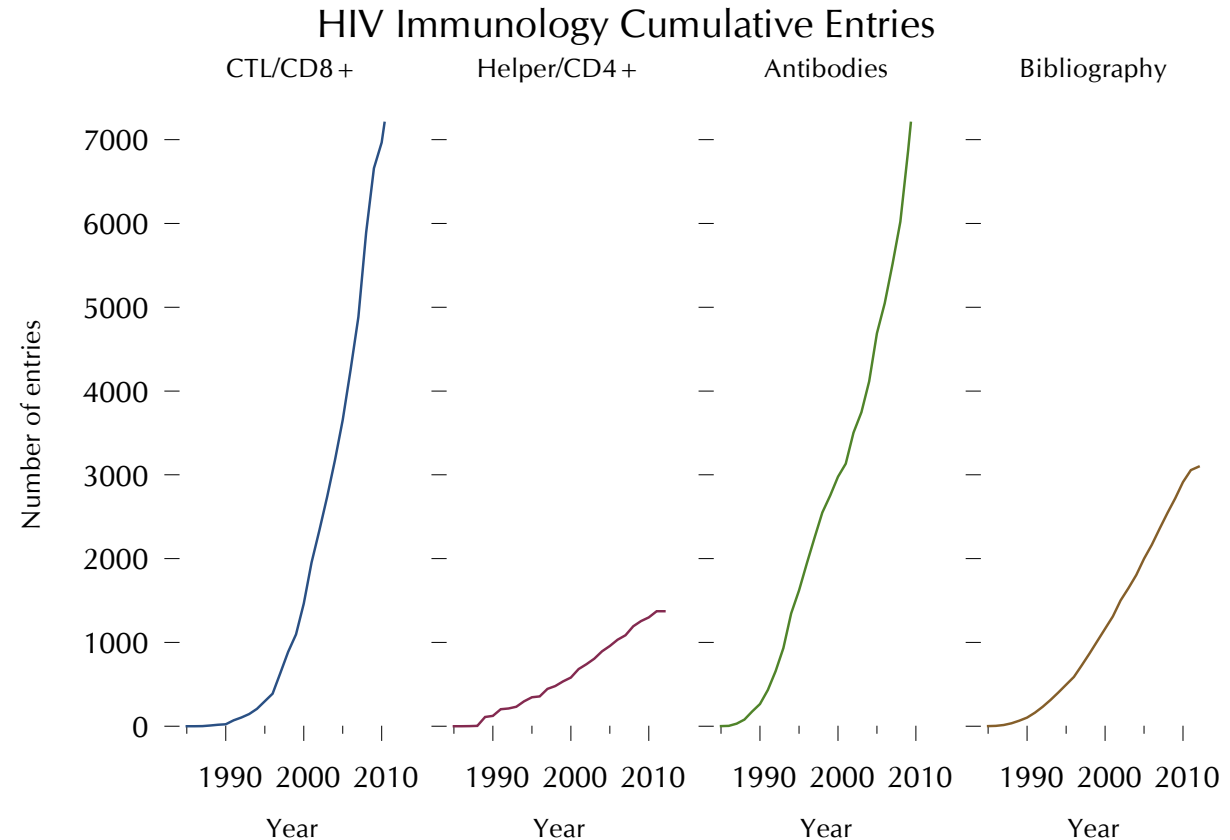
- **Usage:**
  - ~ 70% papers entered in CTL epitope database use HIV Immunology Database resources

Los Alamos
NATIONAL LABORATORY

HIV Molecular Immunology Database

The HIV Molecular Immunology Database is an annotated, searchable collection of HIV-1 cytotoxic and helper T-cell epitopes and antibody binding sites.

Search the Molecular Immunology Database

- CTL/CD8+ Search
- T Helper/CD4+ Search
- Antibody Search
- Search Help

Database Products

- All Database products and publications
- Epitope maps
- Epitope summary tables
- Epitope alignments
- Epitope variants and escape mutations
- The HIV Molecular Immunology Compendium
- About the HIV Molecular Immunology Database
- How to cite this database

Tools and Data Sets

- Tools & Links for immunologists
- HIV "A list" CTL/CD8+ Epitopes (PDF) review article summarizing the best-characterized HIV epitopes
- SIV Epitopes (PDF) review article summarizing known SIV epitopes
- Identifying HLA-Associated Polymorphisms in HIV-1 (PDF) review article summarizing HIV polymorphism associated with escape mutations. Also a table of polymorphisms.
- HLATEM HLA Typing and Epitope Mapping Data Sets
- Standardized Assessments of Neutralizing Antibodies for HIV/AIDS Vaccine Development Assay protocols from Duke Central Reference Laboratory

Los Alamos
NATIONAL LABORATORY

Immunology Database Products

- Epitope maps (species/HLA for T cell epitopes; species/MAb name for Ab)

- Epitope summary tables:
  - All CTL and Helper epitopes and Ab binding sites
  - Variants of CTL epitopes
  - Review of HIV-1 HLA-associated polymorphisms, with a link to a table of polymorphisms, by Zabrina Brumme *et al. (*Under "Tools and Data Sets"*).*
  - Christian Brander keeps an "A list" of HIV CD8+ T-cell epitopes – experimentally validated optimal epitopes with known HLA presenting molecules, will be updated soon
  - "B list" – a comprehensive list of all unique epitopes in the database (unknown HLA, boundaries not fully defined…)
  - All antibodies organized by protein and binding region
  - New: Neutralizing Ab resources: Antibody A-list, Antibody contexts and features

- Tools for immunologists

- Yearly HIV Molecular Immunology Compendium

Los Alamos
NATIONAL LABORATORY

p17 CTL/CD8+ Epitope Map

- Epitopes up to 14 aa long are mapped on HXB2
- HXB2 sequence may differ
- Epitopes with identical boundaries and HLA fields are included in the maps only once
- The epitope maps are interactive!
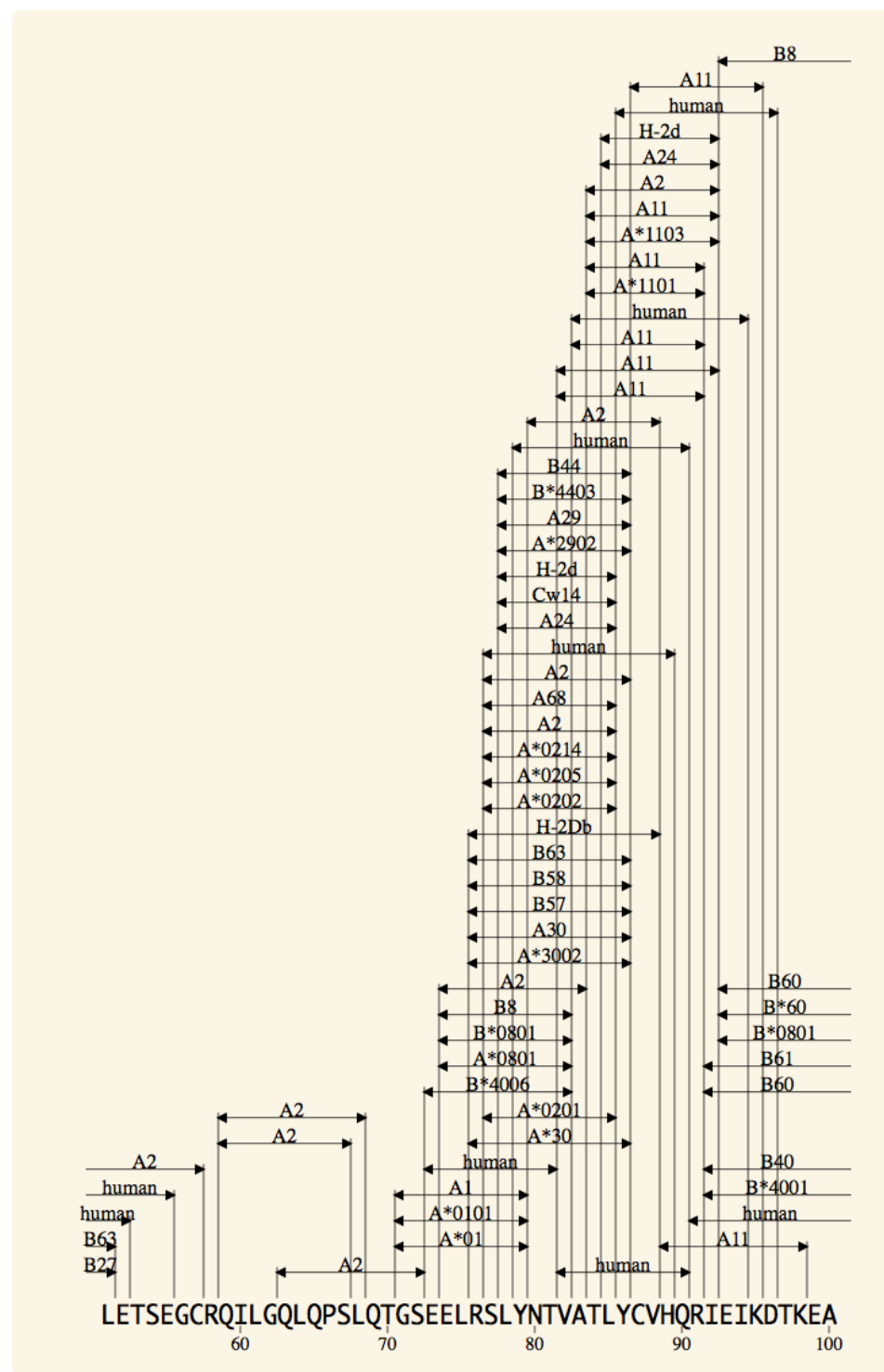
CTL/CD8+ Epitope Summary (B-list)

- List of all epitopes up to 21 aa long
- Unlike epitope maps that show epitope locations, here each epitope sequence is shown

Epitope	Protein	HXB2 Location	Subtype	Species	HLA
MGARASVLSG	p17	1-10	CRF01_AE	human	
ASVLSGGEL	p17	5-13	B	human	
ASILRGGKLDK	p17	5-15	C	human	
SVLSGGQLDR	p17	6-15	B	human	A11
LSGGELDRWEK	p17	8-18		macaque	
GELDRWEKI	p17	11-19	B	human	B*4002, B40
GQLDRWEKI	p17	11-19	B	human	
GKLDSWEKIRLR	p17	11-22	A, CRF01_AE, CRF02_AG	human	
GKLDAWEKIRLR	p17	11-22	CRF01_AE	human	
ELDRWEKIRL	p17	12-21	B, C	human	B63
EKIRLRPGGKKYKL	p17	17-31		human	B27, B7
KIRLRPGGK	p17	18-26	A, A1, B, CRF01_AE	human, transgenic mouse	A*0301, A11, A3, B27, B7
KIRLRPGGKK	p17	18-27	B, C, multiple	human	A*0301, A11, A3, B27
KIRLRPGGKKYKL	p17	18-31		human	A3, B62

Best-defined CTL/CD8+ Epitope Summary (A-list)

- Selective list of best defined epitopes as described by Christian Brander and colleagues

Epitope	Protein	HXB2 Location	Subtype	Species	HLA
GELDRWEKI	p17	11-19		human	B*4002
KIRLRPGGK	p17	18-26		human	A*0301
IRLRPGGKK	p17	19-27	B	human	B*2705
RLRPGGKKK	p17	20-28		human	A*0301
RLRPGGKKKY	p17	20-29	B	human	A*0301
GGKKKYKLK	p17	24-32	B	human	B*0801
KYKLKHIVW	p17	28-36	B	human	A*2402
HLVWASREL	p17	33-41		human	Cw*0804
LVWASRELERF	p17	34-44		human	A30
WASRELERF	p17	36-44	B	human	B*3501
ELRSLYNTV	p17	74-82		human	B*0801
RSLYNTVATLY	p17	76-86	B	human	A*3002, B58, B63
SLYNTVATL	p17	77-85	B	human	A*0201, A*0202, A*0205
LYNTVATL	p17	78-85		human	Cw14
LYNTVATLY	p17	78-86		human	A*2902, B*4403
TLYCVHQK	p17	84-91		human	A*1101
IEIKDTKEAL	p17	92-101		human	B*4001
NSSKVSQNY	p17	124-132	B	human	B*3501

Los Alamos
NATIONAL LABORATORY

Immunology Database: Search

- **T Cells**
  - ☐ Cytotoxic T Lymphocytes (CTL)
  - ☐ Helper T Lymphocytes (T-helper)
  - ☐ Organization is identical for CTL and T-helper
  - ☐ One reference per entry, epitope/HLA combinations are often repeated
- **B Cells (Antibodies)**
  - ☐ One entry for each monoclonal antibodies
  - ☐ Many references per entry (> 400 for some well studied MAbs)

Los Alamos
NATIONAL LABORATORY

CTL/CD8+ T-cell Search

- Can search by HIV protein, Epitope Sequence, Subtype, Immunogen, Vaccine Details, Species, presenting MHC/HLA, Author, Country, Keywords

- Can now search on epitope location and find fuzzy matches, overlaps and embedded epitopes

- Search example:
  - □ SLYNTVATL – 254 entries
  - □ To narrow the search use keyword "escape" – 32 entries

- Additional information provided in the entry:
  - □ Location, Donor MHC/HLA, experimental methods, Notes
  - □ CTL epitope variants if studied in the paper
  - □ Link to all entries for a reference
  - □ PubMed links to papers
  - □ Link to Epitope Maps
  - □ Link to Epitope Alignment (Extracted from HIV-sequence database, includes subtype, country and year of sampling)

• Los Alamos
NATIONAL LABORATORY

CTL/CD8+ T-cell Search

Search CTL/CD8+ T-Cell Epitope Database

Search for

Epitope: ISPRTLNAW

First Author: Pillay

HIV protein	**Proteins with defined epitopes** – ALL – p17 p17-p24 p24 p24-p2p7p1p6	**Proteins with undefined epitopes** – ALL – Gag Gag/Pol Pol Vif
HXB2 location	___ — ___	Results overlap with query location ▾
Epitope	ISPRTLNAW	Results contain query sequence ▾
Epitope name		
Record number		
Subtype	– ALL – ▾	
Immunogen	– ALL – computer prediction HIV-1 and GBV-C co-infection HIV-1 and HCV co-infection HIV-1 exposed seronegative HIV-1 infected monocyte-derived HIV-1 infection	
Vaccine details if Immunogen is Vaccine	Vaccine type	– ALL – ▾
	Vaccine strain	– ALL – ▾
	Vaccine component	– ALL – ▾
	Adjuvant	– ALL – ▾
Species	– ALL – ▾	
MHC/HLA	– ALL – A*01 A*0101 A*02 A*0201 A*02.01 A*020101	
Author	Pillay	☑First ☐Last
Country	– ALL – ▾	
Keywords	– ALL – acute/early infection adjuvant comparison antagonism antibody binding site definition and exposure assay development, comparison, standardization, improvement autologous responses	
Note		

Search   Reset   Click for Search Help

Search CTL/CD8+ T-Cell Epitope Database

Found 1 matching record:

Displaying record number 53832

		Epitope Map

HXB2 Location   p24(15-23)

Author Location   Gag(147-155)   p24 Epitope Map

Epitope   ISPRTLNAW

                                                             Epitope Alignment
                                                             Show epitope
                                                             variants

Subtype   C

Species (MHC/HLA)   human(B57)

Immunogen   HIV-1 infection

Donor MHC/HLA   A*3001, A*66, B*4201, B*5802, Cw*0602, Cw*1701; A*66, A*68, B*57, B*5802, Cw*0602,
                       Cw*0701

Country   South Africa

Experimental
methods   CD8 T-cell Elispot - IFNy

Keywords   epitope processing, responses in children, mother-to-infant transmission, escape,
                       acute/early infection

Notes

- HIV-specific CTLs in infants were shown to be able to select for viral escape variants early in life, despite a lack of escape with the same CTL specificity in the mother. Infant CTL responses may be compromised by transmission of escape variants that arose in the mother and also those that arose in the father, if the father was the source of the mother's infection.
- ISPRTLNAW is the C consensus form of the epitope and was the autologous form in the mother, and was transmitted to her infant. By 33 weeks a new dominant form of the epitope had emerged in the infant, mSPRTLNAW, and two additional variants had arisen, one with a substitution proximal to the epitope, pISPRTLNAW, and lSPRTLNAW.

References

Pillay2005 Thillagavathie Pillay, Hua-Tang Zhang, Jan W. Drijfhout, Nicola Robinson, Helen Brown, Munira Khan, Jagadesa Moodley, Miriam Adhikari, Katja Pfafferott, Margaret E. Feeney, Anne St. John, Edward C. Holmes, Hoosen M. Coovadia, Paul Klenerman, Philip J. R. Goulder, and Rodney E. Phillips. Unique Acquisition of Cytotoxic T-Lymphocyte Escape Mutants in Infant Human Immunodeficiency

Variants details

Displaying record number 53832

HXB2 Location	p24(15-23)	**p24 Epitope Map**
Epitope	ISPRTLNAW	**Epitope Alignment**
	mSPRTLNAW	escape documented in this paper
Variants	lSPRTLNAW	diminished response
	pllSPRTLNAW	not determined
Species (**MHC/HLA**)	human(B57)	

Can go back to epitope entry

Variant Details

Showing all 3 variants.

Variant ID.	1413
Epitope Seq.	ISPRTLNAW
Variant Seq.	mSPRTLNAW
Mutations	I/M
Epitope Location	I1M
HXB2 Location	I15M
Mutation Type	E: escape documented in this paper
Method	CD8 T-cell Elispot - IFNy, Sequence
Note	This is de novo variant seen in infant by week 33 of age. The index peptide was recognized, but not the variant.

Mutation type

Note describing why the variant was designated particular mutation type

Variant ID.	1414
Epitope Seq.	ISPRTLNAW
Variant Seq.	lSPRTLNAW
Mutations	I/L
Epitope Location	I1L
HXB2 Location	I15L
Mutation Type	DR: diminished response
Method	CD8 T-cell Elispot - IFNy, Sequence

Summary table of ~ 2800 variants

CTL/CD8+ Epitope Variant Details

Download CTL/CD8+ epitope variant details as CSV or XLS files.

List of Mutation types

Data last updated at 2013-01-25 11:58:16-07

Epitope ID	Epitope Name	Variant ID	Subtype	Epitope Subtype	Variant Subtype	Protein	HXB2 start	HXB2 end	HLA	Epitope	Variant Epitope	Mutation (epitope)	Mutation (protein)	Mutation Type Code	Mutation Type Description	Methods	Note	References
54532	AI14	1016	B	B	A, M-group	p17	5	19		ASVLSGGELDRWEKI	ASVLSGGkLDaWEKI	R11A, E8K	R15A, E12K	SNSF	subtype-specific non-susceptible form	CD8 T-cell Elispot - IFNy	No cross-recognition of this variant was seen across clades or intra-clade central sequences.	Malhotra2007
54532	AI14	1017	B	B	C	p17	5	19		ASVLSGGELDRWEKI	ASiLrGGkLDkWEKI	R11K, V3I, S5R, E8K	R15K, V7I, S9R, E12K	SNSF	subtype-specific non-susceptible form	CD8 T-cell Elispot - IFNy	No cross-recognition of this variant was seen across clades or intra-clade central sequences.	Malhotra2007
54532	AI14	1018	B	B	B	p17	5	19		ASVLSGGELDRWEKI	ASVLSGGkLDkWEKI	R11K, E8K	R15K, E12K	SNSF	subtype-specific non-susceptible form	CD8 T-cell Elispot - IFNy	No cross-recognition of this variant was seen across clades or intra-clade central sequences.	Malhotra2007
54532	AI14	1019	B	B	B	p17	5	19		ASVLSGGELDRWEKI	ASVLSGGELDkWEKI	R11K	R15K	SNSF	subtype-specific non-susceptible form	CD8 T-cell Elispot - IFNy	No cross-recognition of this variant was seen across clades or intra-clade central sequences.	Malhotra2007
53591	Gag 1.2	54	B		CRF02_AG	p17	8	18		LSGGELDRWEK	LSGGkLDaWEK	E5K, R8A	E12K, R15A	SNSF	subtype-specific non-susceptible form	Intracellular cytokine staining, T-cell Elispot	CRF02 form, LSGGkLDaWEK, does not cross-react with the B clade LSGGELDRWEK elicited response.	Amara2005a
53844	GI9	1569	B			p17	11	19	B40	GELDRWEKI	GELDRWkKI	E7K	E17K	DR, LE	diminished response, literature escape	CD8 T-cell Elispot - IFNy, Sequence	This variant from the HXB2 sequence was present in the restricting HLA-B40-carrying mother, M-1002, but was never detected in her non-HLA-B40-carrying infant, P-1031. Decreased recognition of the E17K variant relative to the index epitope was seen in the mother.	Sanchez-Merino2005
56027	GI9(p17)	1903	B	B	B	p17	11	19		GQLDRWEKI	GeLDRWEKI	Q2E	Q12E	ND	not determined	CD8 T-cell Elispot - IFNy, Sequence	This Asian B Clade optimal epitope differs from the consensus B at one position. It is predicted to be HLA-B40 restricted. Experimentally, B clade consensus peptide was used to challenge CTL response in subjects commonly carrying the Asian B-type epitope.	Zhai2008
55632		11	A, CRF02_AG, CRF01_AE	A, AG	AE	p17	11	22		GKLDSWEKIRLR	GKLDaWEKIRLR	S5A	S15A	SSF	subtype-specific susceptible form	CD8 T-cell Elispot - IFNy	1 subject responded to peptide GKLDSWEKIRLR from subtypes CRF02_AG and A and to peptide GKLDaWEKIRLR from subtype CRF01_AE.	Aidoo2008
54629	GAG-03	1957	B	B	C	p17	17	34		EKIRLRPGGKKKYRLKHL	EKIRLRPGGKKhYmLKHL	K12H, R14M	K28H, R30M	SSF	subtype-specific susceptible form	CD8 T-cell Elispot - IFNy, Sequence	This Clade C consensus synthetic peptide variant from an immunodominant region, differs from the immunodominant Clade B consensus at 2 amino acids (11.1%) and both were recognized by subtype-B-infected subjects.	Zhao2007
53201	KK9	31	B			p17	18	26	A3	KIRLRPGGK	KIRLRPGGq	K9Q	K26Q	E, P	escape documented in this paper, processing	CD8 T-cell Elispot - IFNy, Flow cytometric T-cell cytokine assay	Variant inhibits processing, resulting in rapid decline in the KK9 specific CD8+ T-cell response.	Allen2004
55770	KK9	153	B			p17	18	26	A3	KIRLRPGGK	KIRLRPGGr	K9R	K26R	SF	susceptible form	Flow cytometric T-cell cytokine assay	KIRLRPGGK was recognized by 3 patients. The autologous sequence in one patient was KIRLRPGGr which induced high frequency response.	Daucher2008
55233		790	B, CRF01_AE	B		p17	18	26	A3	KIRLRPGGK	KIRLRPGGr	K9R	K26R	IE	inferred escape	Sequence	Patient was superinfected with three strains, B1, B2 and CRF01_AE. This variant developed in B1 to include 42% of the viruses within 4 years.	Kozaczynska2007

Under database Products
(http://www.hiv.lanl.gov/content/immunology/index.html)

Go to "Epitope variants and escape mutations"

Los Alamos
NATIONAL LABORATORY

Antibody Search

- Can search by
  - HIV protein, Epitope Sequence, Subtype, Immunogen, Vaccine Details, Species, presenting MHC/HLA, Author, Country, Keywords
  - MAb ID (Ab lists by name and by binding type are provided)
  - Ab type (by binding site, for example binding to similar region like V3 or near a common functional domain like CD4 binding site CD4Bs)
  - Isotype
- Search examples:
  - 2F5 – 1 record with 463 references
  - Ab type: gp120 CD4BS – 200 records

Antibody Search

Search Antibody Database

	Proteins with **defined epitopes**	Proteins with **undefined epitopes**	
HIV protein	– ALL – p17 p17–p24 p24 p24–p2p7p1p6	– ALL – p24 Gag RT Pol	

HXB2 location	[ ] – [ ]	Results overlap with query location ▼
Epitope	[ ]	Results contain query sequence ▼
Record number	[ ]	
MAb ID	[ ]	(List by name) (List by type)
Subtype	– ALL – ▼	
Immunogen	– ALL – anti–idiotype autoimmune disease HIV–1 exposed seronegative HIV–1 infection HIV–2 infection in vitro stimulation or selection	

Vaccine details if Immunogen is Vaccine	Vaccine type	– ALL – ▼
	Vaccine strain	– ALL – ▼
	Vaccine component	– ALL – ▼
	Adjuvant	– ALL – ▼

Ab Type	– ALL – C-domain C-HR C-term Env oligomer flap region gp120 adjacent to CD4BS	
Species	– ALL – ▼	
Isotype	– ALL – IgA IgA1 IgA2 IgA22a IgE IgG	
Author	[ ]	Search only for ☐First ☐Last author ◉ Show only this author's references ○ Show all references
Country	– ALL – ▼	
Keywords	– ALL – acute/early infection ADCC adjuvant comparison antibody binding site definition and exposure antibody generation antibody interactions	◉ Show only notes containing selected keyword(s) ○ Show all notes
Note	[ ]	◉ Show only notes matching this text ○ Show all notes

[Search] [Reset] Click for Search Help

Can search by HXB2 location,
Find overlaps, fuzzy matches
Embedded epitopes

Can show only notes and
references containing
selected keywords or user's
text (as apposed to showing
matching Ab entries with all
notes)

Los Alamos
NATIONAL LABORATORY

Antibody Search

Displaying record number 815

MAb ID	2F5 (IAM 2F5, IAM-41-2F5, IAM2F5, c2F5)	
HXB2 Location	gp160(662-667)	**gp160 Epitope Map**
Author Location	gp41(662-667 BH10)	
Research Contact	Hermann Katinger, Institute of Applied Microbiology, Vienna, or Polymun Scientific Inc., Vienna, Austria	
Epitope	ELDKWA	**Epitope Alignment**
Ab Type	gp41 adjacent to cluster II, C-term, gp41 MPER (membrane proximal external region)	
Neutralizing	L P	
Species (Isotype)	human(IgG3κ)	
Immunogen	HIV-1 infection	
Keywords	acute/early infection, adjuvant comparison, anti-idiotype, antibody binding site definition and exposure, antibody generation, antibody interactions, antibody sequence variable domain, assay development, standardization and improvement, autoantibody or autoimmunity, autologous responses, binding affinity, brain/CSF, co-receptor, complement, dendritic cells, drug resistance, enhancing activity, escape, genital and mucosal immunity, HAART, ART, HIV exposed persistently seronegative (HEPS), immunoprophylaxis, immunotherapy, immunotoxin, isotype switch, kinetics, mimics, mimotopes, mother-to-infant transmission, neutralization, rate of progression, responses in children, review, SIV, structure, subtype comparisons, supervised treatment interruptions (STI), therapeutic vaccine, vaccine antigen design, vaccine-induced immune responses, variant cross-recognition or cross-neutralization, viral fitness and reversion	

Notes

- 2F5: 2F5 neutralized infection of PBLs with various HIV-1 strains with high potency. However, 2F5 did not inhibit transcytosis of cell-free or cell-associated virus across a monolayer of epithelial cells. A mixture of 13 MAbs directed to well-defined epitopes of the HIV-1 envelope, including 2F5, did not inhibit HIV-1 transcytosis, indicating that envelope epitopes involved in neutralization are not involved in mediating HIV-1 transcytosis. When the mixture of 13 MAbs and HIV-1 was incubated with polyclonal anti-human γ chain, the transcytosis was partially inhibited, indicating that agglutination of viral particles at the apical surface of cells may be critical for HIV transcytosis inhibition by HIV-specific Abs. *Chomont2008* (**neutralization**)
- 2F5: The lipid binding properties of 2F5, and the similarity to binding properties of anti-lipid mAbs, are discussed. Potential role of liposomes containing lipid A for induction of NAbs to lipids of HIV-1 is reviewed. *Alving2008* (**autoantibody or autoimmunity, review**)
- 2F5: A reference panel of recently transmitted Tier 2 HIV-1 subtype B envelope viruses was developed representing a broad spectrum of genetic diversity and neutralization sensitivity. The panel includes viruses derived from male-to-male, female-to-male, and male-to-female sexual transmissions, and CCR5 as well as CXCR4 using viruses. The envelopes displayed varying degrees of neutralization sensitivity to 2F5, with 14 of 19 enevlopes sensitive to neutralization by this Ab. *Schweighardt2007* (**assay development, standardization and improvement, neutralization**)
- 2F5: This review summarizes data on possible vaccine targets for elicitation of neutralizing Abs and discusses whether it is more practical to design

Neutralizing Antibody Resources

Summary table

A table presenting most broadly neutralizing HIV-1 antibodies, with links to papers, Ab sequences and structure, notes on breadth of neutralization, where to find Ab contacts or key residues, and heavy and light chain composition.

- Best neutralizing antibodies

Search interface

This interface allows you to search for exact coordinates of important neutralizing antibody binding sites and other HIV-1 Env features.

- Neutralizing antibody contexts and features

Spreadsheet

A summary of the information from the search interface above, presented in a single .xls spreadsheet. Each row of the table corresponds to one residue of HIV-1 Env, and each column represents a protein feature or set of known binding residues of broadly neutralizing antibodies. Loops and other features of Env are shown in the first 3 columns on the left. The entropy (sequence variability) of each residue is presented numerically and color coded. Abbreviated references are listed under each column heading, and full references are on the second page of the Excel file.

- Spreadsheet of neutralizing antibody contexts and features (.xls)

Los Alamos
NATIONAL LABORATORY

Summary of Best Neutralizing Antibodies

Download summary of best neutralizing antibodies as CSV or XLS files.

This is a list of most broadly neutralizing HIV antibodies, with links to papers, Ab sequences and structure, notes on breadth of neutralization, Ab contact or key residues and heavy and light chain composition.
Note: this is a work in progress, so not all relevant papers and antibodies are listed.

Mab	Binding site	Author Journal Pmid	First paper	Breadth of neutralization with IC50<50 µg/ml	Breadth of neutralization with IC80 or IC90<50 µg/ml	Structure, PDB ID	Ab sequence	Heavy chain	Light chain	Germline Ab sequence	Ab binding affinity	Listings of antibody contact or key residues
VRC01	CD4bs	Wu2010 Science 20616233	YES	91% of 190 isolates, representing major HIV-1 clades	86% of 190 isolates, representing major HIV-1 clades, with IC80		GI:294875838 -- heavy chain variable region GI:294875848 -- light chain variable region	V: IGHV1-02*02 D: IGHD3-16*01 (or *02) J: IGHJ1*01 or IGHJ2*01	V: IGKV3-11*01 J: IGKJ2*01	Fig. S5	Bound strongly to RSC3 and gp120 and weakly to ΔRSC3, Fig. 2 and S4.	
		Zhou2010 Science 20616231				3NGB				Fig. S12	Figs. 5, 6, S3	Env, defined by crystal structure: Fig S1. Antibody, defined by crystal structure: Fig. S9
		Wu2011 Science 21835983								Sequence, Figs. 1, S14, S18. Phylogenic analysis, Fig. 5, Fig. S13		Antibody, defined by crystal structure, compared to key residues of other CD4bs antibodies, Fig. S4.
		Scheid2011 Science 21764753		100% of 118 isolates representing major HIV-1 clades							Fig. 3, Table S9.	Antibody, defined by crystal structure in Zhou2010, Fig. 4, Fig. S3, and Fig. S4 provide comparisons with other CD4bs Nabs.
		Walker2011 Nature 21849977		93% of 162 isolates representing major HIV clades	89% of 162 isolates representing major HIV clades, with IC90							

http://hiv-dev.lanl.gov:8081/content/immunology/tables/tables.html
Epitope Summary Tables
Best Neutralizing Antibodies

Los Alamos NATIONAL LABORATORY

Spreadsheet of Ab contexts and features

Regions_1: signal peptide, cleavage sites, disulfied bonds, hypervariable regions, integrin α4β7 binding, Lectin DC-sign binding, gp41 regions: Leucine/Isoleucine	Regions_2: Variable loops, gp41 regions: Kennedy epitope, GxxxG motif, Cytoplasmic tail LLP1, LLP2, LLP3; RxKR motif	Regions_3: gp120 Regions related to the CD4 binding site, gp41 regions: fusion peptide, Leucine/Isoleucine zipper, fusion domain, YXXL motif	Entropy score	HXB2 position	HXB2 AA	CH103 contact & bonding	b12 contacts	17b contacts	PG9 contacts PG9-related class: PG16, PGT141, 145, CH01-CH04	PG9 sensitivity PG9-related class: PG16, PGT141, 145, CH01-CH04	10E8	2F5	4E10	Z13
Cys 131 linked to Cys 157	forms V1 loop		0.02	157	C									
	V2		0.30	158	S									
	V2		0.17	159	F									
	V2		0.40	160	N				PG9 glycan cont	Loss of glycan confers resistance				
	V2		1.57	161	I									
	V2		0.29	162	S									
	V2		0.36	163	T									
	V2		1.66	164	S									
	V2		1.32	165	I				PG9 contact					
	V2		1.13	166	R									
	V2		0.60	167	G				PG9 contact					
	V2		0.55	168	K				PG9 contact					
	V2		1.97	169	V				PG9 contact	169/171 neg charge (E) resistant, pos charge (K) sensitive				
	V2		1.46	170	Q				PG9 contact					
	V2		1.24	171	K				PG9 contact	169/171 neg charge (E) resistant, pos charge (K) sensitive				
	V2		1.51	172	E									
	V2		1.17	173	Y				PG9 contact	Y increases sensitivity				
	V2		0.42	174	A									
	V2		0.50	175	F									
	V2		0.17	176	F									
	V2		0.35	177	Y									
	V2		1.23	178	K									
LDI/LDV tripeptide in V:	V2		0.96	179	L									
LDI/LDV tripeptide in V:	V2		0.03	180	D	mutation may induce a conformational change that exposes epitopes in both the gp120 and the gp41 portions of the envelope protein,								
LDI/LDV tripeptide in V:	V2		0.97	181	I									
	V2		0.69	182	I									
	V2		1.18	183	P									
	V2		1.06	184	I									
V2-hypervariable	V2		1.70	185	D									
V2-hypervariable	V2		1.87	186	N									
V2-hypervariable	V2		1.82	187	D									
V2-hypervariable	V2		1.87	188	T									
V2-hypervariable	V2		1.81	189	T									
V2-hypervariable	V2		1.98	190	S									
	V2		0.16	191	Y									
	V2		0.85	192	K									
	V2		0.14	193	L									
	V2		0.50	194	T									
	V2		0.90	195	S									
Cys 126 linked to Cys 196,	V2 loop end		0.04	196	C									
			0.11	197	N									
			0.46	198	T									
			0.09	199	S	4, VRC01, CD4-IgG b12								
			1.30	200	V									
			0.59	201	I									

Los Alamos
NATIONAL LABORATORY

Tools for Immunologists

- **Sequence Locator** Finds epitope location on the reference genome
- **QuickAlign** Aligns amino acids or nucleotides against our alignments
- **PeptGen** Generates overlapping peptides for any protein
- **PepMap** Generates peptide maps in Fasta, HTML and PDF formats
- **ELF** Epitope Location Finder
- **N-Glycosite** Finds N-linked glycosylation sites
- **Mosaic** Generates candidate vaccine protein cocktails
- **Motifscan** Scans alignments for HLA binding motifs
- **Highlighter** Highlights matches and mismatches in a set of aligned sequences
- **Heatmap** Displays and organizes neutralization or other quantitative data.
- And more …

Los Alamos
NATIONAL LABORATORY

HIV Molecular Immunology Database: Tools & Links

Tools Produced by the Los Alamos HIV Databases

- QuickAlign Align amino acids or nucleotides against our alignments
    - Epilign and PrimAlign have been replaced by QuickAlign
- PeptGen Generate overlapping peptides for any protein
- PepMap Generate peptide maps in Fasta, HTML and PDF formats
- Hepitope Search for hopeful epitopes based on HLA enrichment
- HLA Frequency Analysis Tools Calculate HLA frequencies or HLA linkage disequilibrium in a population
- ELF Epitope location finder
- Motif Scan Scan alignments for HLA binding motifs
    - HLA genotype/serotype dictionary
    - HLA genotype/motif dictionary
    - HLA supertype dictionaries
- HIV/SIV Sequence Locator Tool Formerly the *HXB2 Numbering Engine*
- SeqPublish Produce pretty alignments for publication
- BLAST Search sequences against our annotated HIV sequences
- ODprep/ODfit Calculate antibody titers based on concentration and optical density data
- Heatmap Display a table of numbers using colors to represent the numerical values
- Mosaic Generate candidate vaccine protein cocktails
    - Epicover Epitope coverage assessment tool
    - Posicover Positional epitope coverage assessment tool
- N-Glycosite Find N-linked glycosylation sites
- Highlighter Highlight matches and mismatches in a set of aligned sequences

- All Tools List of all software and tools in **both** the HIV sequence and immunology databases

External Tools for Epitope Prediction

- BIMAS HLA Peptide Binding Predictions Ranks potential n-mer peptides based on a predicted half-time of dissociation to HLA class I molecules
- SYFPEITHI Epitope Prediction Predicts the binding of a given amino acid sequence to a defined HLA type
- PAProC Predicts cleavages by human and yeast 20S proteasomes
- PREDEP MHC class I epitope prediction
- MHCPred Predicts MHC/peptide or TAP/peptide IC_{50} binding values
- Microsoft Research Epitope Predictor Computes the probability that a given n-mer is a T-cell epitope restricted to a given HLA allele

Heatmap

- A heatmap is a graphical way of displaying a table of numbers by using colors to represent the numerical values.

- This is a very nice way to display large sets immunological data (eg antibody and target data), statistically clustering "like" behaviors

- Commonly used in gene expression array data

- We have two options in the immunology database:
  - Hierarchical clustering: We use R code for this, and if you are comfortable coding in R you can better manipulate fine details
  - An in-house developed k-means clustering strategy that incorporates an error model

Heatmap   Hierarchical Clustering

Purpose: A heatmap is a graphical way of displaying a table of numbers by using colors to represent the numerical values. The clustering algorithm groups related rows and/or columns together by similarity. For details see Heatmap Hierarchical Explanation.

Input

Paste your data here

[Sample Input]

Or upload a data file [   ] [Browse...]

Options

Use log transformation? ○ Natural ● Base10 ○ No transformation

Usually use log base 10 for antibody data

Use threshold value? ● No ○ Below threshold value: [   ]   ○ Above threshold value: [   ]

Antibody data is often censored

Cluster method [Complete ▾]   Distance method [Euclidean ▾]

Heatmap ☑ Cluster by ○ Rows only ○ Columns only ● Both ○ None

Colors   Palette: [Brewer ▾]

Higher values: ● High intensity color ○ Low intensity color

NAb sera are dilutions, so high values are potent

Number of colors: [9]

Color key ranges: [   ]

Margins   Bottom Margin: [5]   Right Margin: [12]
Label sizes Row: [100% ▾]   Column: [100% ▾]

MAbs are concentrations, so low values are potent

See the stability of clusters using bootstraps ☐ Iterations [100 ▾]

By ☑ Standard Bootstrap (slow but accurate)

☑ scale bar
show only labels with >= [50 ▾] % bootstrap support

☐ pvclust (fast but approximate)

[Submit] [Reset]

Los Alamos
NATIONAL LABORATORY

K-means clustering, k = 3
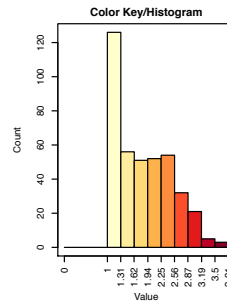
Three clusters, red, blue, yellow

Bootstrap resampling to explore robustness of clusters given sampling limitations

Can add an error model to explore the Impact of experimental reproducibility

Robust blue cluster
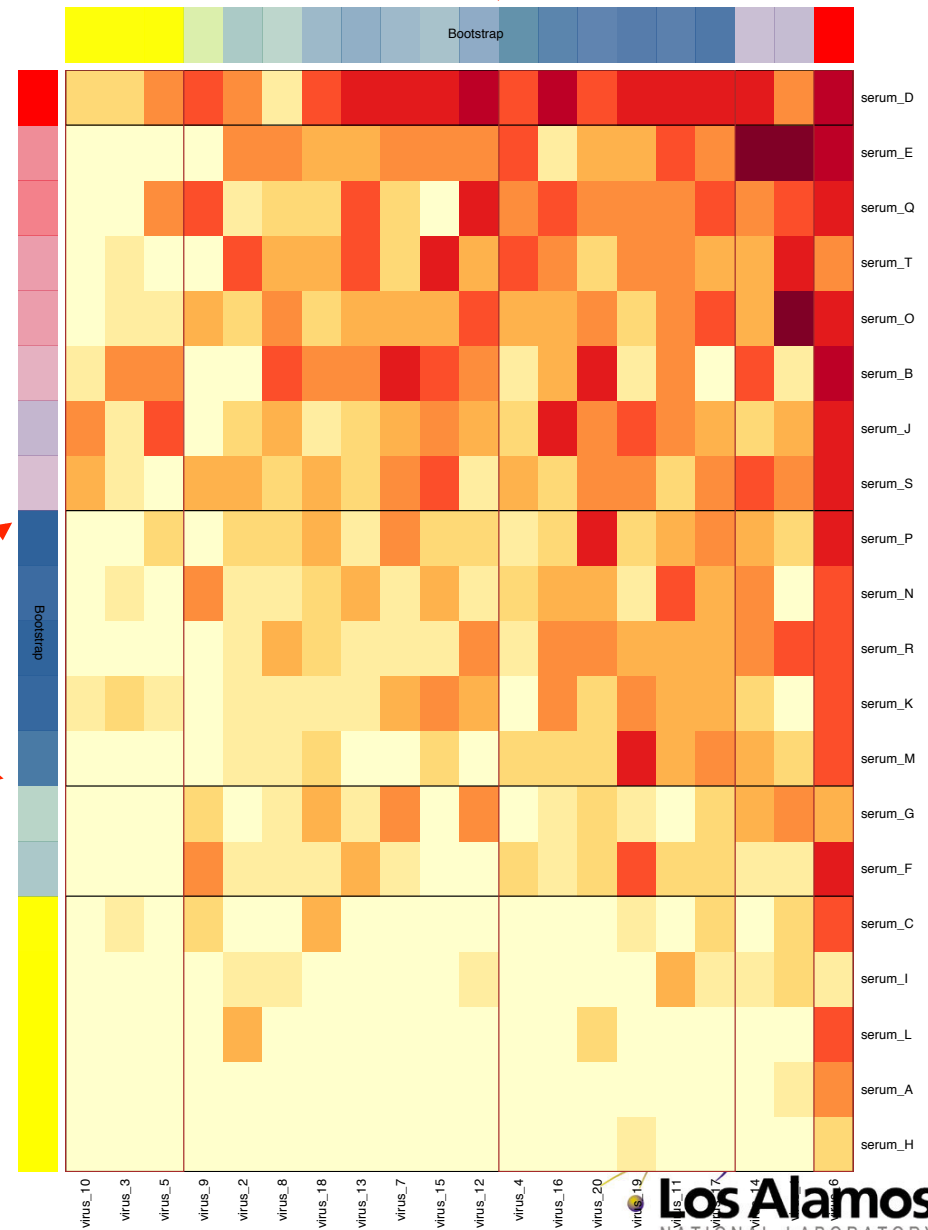
QuickAlign

- Generates an alignment of your HIV-1 amino acid or nucleotide sequence against our web alignments

- Can be used to align epitopes, functional domains, or any protein or nucleotide region of interest

- Calculates frequency of variants to the query sequence and summarizes both by subtype and all subtypes together

- Calculates frequency of amino acid or nucleotide by position and summarizes both by subtype and all subtypes together

Los Alamos
NATIONAL LABORATORY

QuickAlign

formerly Epilign and Primalign

Purpose: Align a desired region from our Web alignments, with or without user-provided sequence(s). Details below.

Retrieve alignment(s) based on sequence

Paste your sequence(s) here
[Sample Input]

or upload sequence file   [   ] (Browse...)

-- OR leave both fields above blank, and --

Retrieve alignment(s) based on coordinates

Sequence coordinates   [1]   start [end]   end

Gene/region/protein   (Complete ⬍)

Options

Organism   ⦿ HIV1   ○ HIV2   ○ SIV

Sequence type   ○ nucleotide   ○ protein   ⦿ let program decide

Alignment type to use   ('Web' alignment (all complete sequences) ⬍)

Delete Gaps and shift sequence toward C-terminus   ○ yes ⦿ no
(protein only)

Display wide output   ○ yes ⦿ no

Calculate frequency by position   ☑   cut-off [95] %

Include surrounding region   ☐

(Submit)   (Reset)

QuickAlign: example of output

- **Query peptide**: SLYNTVATL

- Sequence names include subtype, country and year of sampling

- Identical sequences are shown in red

Query:	SLYNTVATL
Query Length:	9
HXB2 Location:	Gag 77-85 = p17 77-85
Alignment:	GAG, 458 sequences

Summarize

```
Query                SLYNTVATL
A1.KE.86.ML170       --F------
A1.KE.94.Q23         --F------
A1.SE.94.SE7253      --F----V-
A1.SE.94.SE7535      ---------
A1.SE.95.SE8538      ---------
A1.SE.95.SE8891      ---------
A1.SE.95.UGSE8131    ---------
A1.TZ.97.97TZ03      --F----V-
```

Los Alamos
NATIONAL LABORATORY

QuickAlign: sequence variant summary

Summary for subtype A

Variant	Count	Percent
SLYNTVATL		
--F------	11	47.83
---------	7	30.43
--F--I-V-	1	4.35
--F----V-	1	4.35
-------V-	1	4.35
----L----	1	4.35
--F-A--V-	1	4.35

Total sequences = 23
Number of variants = 7

Mutation percentages
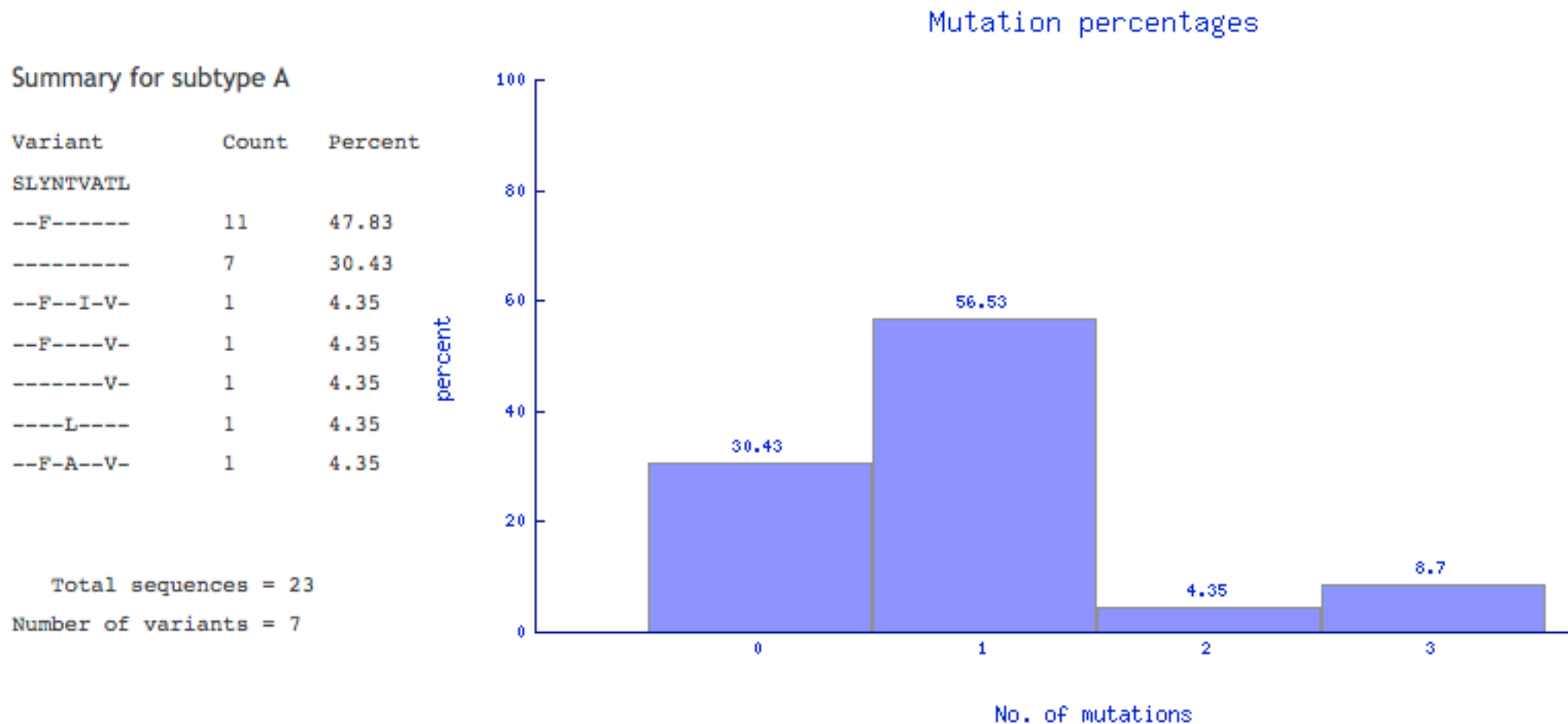
- Variant frequency summary by subtype and all subtype together

QuickAlign:
Frequency by position

See full raw counts

cutoff: 95%

Position	Percentage and raw count of non-gap			Non-gap/total (percentage)
1	S: 99.90% (3113)	other: 0.10% (3)		3116/3119 (100.00%)
2	L: 98.90% (3068)	other: 1.10% (34)		3102/3119 (99.55%)
3	Y: 52.71% (1633)	F: 43.77% (1356)	other: 3.52% (109)	3098/3119 (99.42%)
4	N: 99.68% (3104)	other: 0.32% (10)		3114/3119 (99.94%)
5	T: 92.86% (2887)	A: 5.05% (157)	other: 2.09% (65)	3109/3119 (99.78%)
6	V: 79.35% (2448)	I: 18.15% (560)	other: 2.50% (77)	3085/3119 (99.01%)
7	A: 92.95% (2889)	V: 6.53% (203)	other: 0.51% (16)	3108/3119 (99.74%)
8	T: 72.52% (2254)	V: 27.06% (841)	other: 0.42% (13)	3108/3119 (99.74%)
9	L: 99.00% (3078)	other: 1.00% (31)		3109/3119 (99.78%)

Sequence logo (Subtype: All)
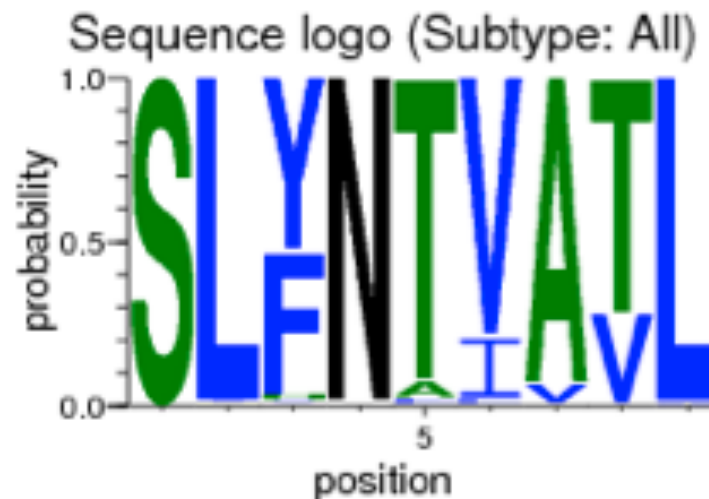
PeptGen

- Creates maps of overlapping peptides on proteins to aid in peptide design for mapping epitopes
- Consensus sequences for all HIV subtypes for all proteins are available
- Use alignments to design comparable sets of peptides (for example, to compare clades)
- INPUT
  - ☐ Query sequence or aligned sequences
  - ☐ Desired length of peptides, peptide overlap,
  - ☐ Forbidden C- and N-terminal amino acids
- OUTPUT
  - ☐ Maps of overlapping peptides (forbidden AAs are taken into account)
  - ☐ Simplified output for ordering
  - ☐ Highlighted forbidden amino acids
  - ☐ Hydrophobicity scores for the peptides are available
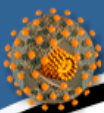
PeptGen Peptide Generator

Purpose: Given an amino acid sequence, this tool generates and displays sets of overlapping peptides that can be used for peptide design and epitope mapping.

How to use: Paste amino acid sequence(s) in any valid format into the window below, or upload a file of sequences. Change the default values for the other parameters as needed and hit Submit. You can also use an HIV consensus sequence as input.

For additional details, see PeptGen Explanation.

Input

Paste your input here
[SAMPLE INPUT - single sequence]
[SAMPLE INPUT - alignment]

or upload your file   [   ] [Browse...]

If you are submitting an alignment check this box   ☐

Options

Make peptides of length	18	C-terminal forbidden amino acids	[ ]
Overlap peptides by	10	N-terminal forbidden amino acids	[ ]
Shorten by	3	Apply proline rule	⦿ Yes ◯ No
Lengthen by	2	Calculate hydropathy	⦿ Yes ◯ No
		Color amino acids	⦿ Yes ◯ No

Output

Produce simple output   ☐

Duplicate peptides   ⦿ flag ◯ remove

Gaps   ⦿ remove ◯ save

Output style   ⦿ classic ◯ new

[Submit] [Reset]

PeptGen: output

Classic output:

PeptGen Results

Download a copy of these results in format: (Text) (PostScript) (PDF)

```
                            Word length: 18
        Overlap consecutive peptides by: 10
                             Shorten by:
                             Lengthen by:
           Forbidden C-term amino acids:
           Forbidden N-term amino acids:
           Number of peptides generated: 9
                          Sequence names: CON_B
                                          CON_C
                                          CON_G


HIVWASRELERFAVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEVKDTKEALEKIEEEQNKSK
HIVWASRELERFAVNPGL (18)
-L-----------L-----
-L-----------L--D-
         LERFAVNPGLLETSEGCR (18)
         -----L-----------K
         -----L--D----A---Q
                  GLLETSEGCRQILGQLQP (18)
                  ---------K--IK----
                  D----A---Q--M-----
                           CRQILGQLQPSLQTGSEE (18)
                           -K--IK----A----T--
                           -Q--M-----A----T--
                                    QPSLQTGSEELRSLYNTV (18)
                                    --A----T----------
                                    --A----T------F---
                                             EELRSLYNTVATLYCVHQ (18)
                                             ----------------E
                                             ------F-----------
                                                      TVATLYCVHQRIEVKDTK (18)
                                                      ---------EK---R---
                                                      ------------------
                                                               HQRIEVKDTKEALEKIEE (18)
                                                               -EK---R------D----
                                                               --------------EV-K
                                                                        TKEALEKIEEEQNKSK (16)
                                                                        -----D---------Q
                                                                        ------EV-KI-K--Q
```

Simple output, "new"

```
13  QPSLQTGSEELRSLYNTV 5_1.1
14  QPALQTGTEELRSLYNTV 5_2.1
15  QPALQTGTEELRSLFNTV 5_3.1

16  EELRSLYNTVATLYCVHQ 6_1.1
17  EELRSLYNTVATLYCVHE 6_2.1
18  EELRSLFNTVATLYCVHQ 6_3.1

19  TVATLYCVHQRIEVKDTK 7_1&3.2
20  TVATLYCVHEKIEVRDTK 7_2.1

21  HQRIEVKDTKEALEKIEE 8_1.1
22  HEKIEVRDTKEALDKIEE 8_2.1
23  HQRIEVKDTKEALEEVEK 8_3.1

24  TKEALEKIEEEQNKSK 9_1.1
25  TKEALDKIEEEQNKSQ 9_2.1
26  TKEALEEVEKIQKKSQ 9_3.1
```

Distinct

Duplicate

Los Alamos
NATIONAL LABORATORY

N-Glycosite

- Highlights and tallies predicted N-linked glycosylation sites (Nx[ST] patterns, where x can be any amino acid)

- NP[ST] pattern can be excluded

```
                                    210        220        230        240        250        260        270
                                     |          |          |          |          |          |          |
A1.KE.93.Q23-17      QACPKVSFEP IPIHYCTPAG FAILKCKDEG FNGTGL--CK NVSTVQCTHG IKPVVSTQLL LNGSLAEKNI
B.FR.HXB2            QACPKVSFEP IPIHYCAPAG FAILKCNNKT FNGTGP--CT NVSTVQCTHG IRPVVSTQLL LNGSLAEEEV
C.BR.92.92BR025     QACPKVSFDP IPIHYCAPAG YAILKCNNKT FNGTGP--CN NVSTIQCTHG TKPVVSTQLL LNGSLAEEEI
D.UG.94.94UG1141    QACPKMTFEP IPIHYCAPAG FAILKCNEKK FNGTGP--CK NVSTVQCTHG IKPVVSTQLL LNGSLAEEEI
01_AE.CF.90.90CF11697 QACPKVTFDP IPIHYCTPAG YAILKCNEKN FNGTGP--CK NVSSVQCTHG IKPVVSTQLL LNGSLAEEDI
02_AG.CM.97.97CM-MP807 QACPKVSFEP IPIHFCAPAG FAILKCKDKE FNGTGP--CK NVSTVQCTHG IKPVVSTQLL LNGSLAEEKV
CPZ.CM.-.CAM3       QACPKTSFEP IPIHYCATPG YAIMKCNMPN FNGTGTGRCN NISTVQCTHG IRPVVTTQLI LNGSVAENKT
O.CM.-.ANT70        QACPKVSFEP IPIHYCAPAG YAIFKCNSTE FNGTGT--CR NITVVTCTHG IRPTVSTQLI LNGTLSKGKI
```

N-glycosylation Sites In Each Submitted Suquence

Sequence Name	N-glycosylation Sites Numbers
A1.KE.93.Q23-17	23
B.FR.HXB2	24
C.BR.92.92BR025	23
D.UG.94.94UG1141	24
01_AE.CF.90.90CF11697	24
02_AG.CM.97.97CM-MP807	24
CPZ.CM.-.CAM3	25
O.CM.-.ANT70	25

A total of 192 N-glycosylation sites in 8 sequences have been found.

The total N-Linked glycosylation sites is 192. (Click the numbers and see the details)

- View the N-Linked glycosylation sites in NX[ST]pattern.
  - The single N-linked glycosylation site count is: 192.
    - The NXS combination count is: 62.
    - The NXT combination count is: 130.
  - The contiguous N-linked glycosylation site (NN[ST][ST]) count is: 0.

N-Glycosite (continuation)

■ Tallies number of N-linked glycosylation sites per alignment position and displays as a downloadable table

Number of N-glycosylation Sites by Position

Download tab-delimited file

Position	31	60	69	102	105	109	110	111	112	115	122	128	145	149	174	179	182	184	185	195	227
A1.KE.93.Q23-17		N			N				N				N	N						N	
B.FR.HXB2		N					N				N		N	N		N				N	
C.BR.92.92BR025		N			N						N		N	N		N		N		N	
D.UG.94.94UG1141		N		N							N		N	N		N			N	N	
01_AE.CF.90.90CF11697		N						N			N	N	N	N		N			N	N	
02_AG.CM.97.97CM-MP807		N				N				N	N		N	N					N	N	
CPZ.CM.-.CAM3		N	N					N			N		N	N		N				N	
O.CM.-.ANT70	N	N												N	N		N			N	N
Total	1	8	1	1	2	1	1	2	1	1	6	1	7	8	1	5	1	1	3	8	1

Los Alamos
NATIONAL LABORATORY

ELF

- If you have a peptide that reacts with CD8+ T cells from a person with known HLA type, enter:
  - ☐ The peptide that reacts with CD8+ T-cells
  - ☐ The HLA type of the person with the reactive CD8+ T cells

- ELF will help identify the possibly reactive epitope by
  - ☐ Highlighting appropriate HLA anchor motifs in the peptide
  - ☐ Aligning all known epitopes embedded in the peptide from the database to your query sequence, with links to epitope entries
  - ☐ Finding potential epitopes based on Immune Epitope Database (IEDB) binding predictions http://www.immuneepitope.org/
- Other useful information provided:
  - ☐ Genomic location of your peptide
  - ☐ Database records for known CTL epitopes in this region, regardless of HLA.

ELF

Epitope Location Finder

Purpose: search a submitted protein sequence for (1) known epitopes from our immunology databases, (2) epitopes predicted by consensus binding motifs, and (3) epitopes predicted by the IEDB binding algorithm. For details see ELF Explanation.

Input

Paste protein sequence	DTVLEDMNLPGRWKPKMIG	<50 amino acids, raw format

Options

Show known epitopes ☑ from CTL and Helper databases

Find potential epitopes ☑ based on anchor residues

Choose HLA(s)
(Class I and Class II)

Use control-click for multiple selection

By genotype
```
A*3004
A*3101
A*3201
A*3303
A*6601
A*6801
A*6802
```

By serotype
```
A33(19)
A69(28)
A68(28)
A30(19)
A66(10)
A1
A2
```

HLA selection is synchronized between 2 analysis options

Find potential epitopes ☑ based on IEDB binding predictions

Choose HLA(s) or MHC(s)
(synchronized with genotype selections above)

HLA Class I
```
A*6611
A*6612
A*6613
A*6614
A*6615
A*6801
A*6802
```

HLA Class II
```
DRB3*0224
DRB3*0225
DRB3*0301
DRB3*0303
DRB4*0101
DRB4*0103
DRB5*0101
```

Animal MHC Class I
```
chimpanzee
  Patr-A*0101
  Patr-A*0201
  Patr-A*0301
  Patr-A*0302
  Patr-A*0401
  Patr-A*0402
```

Animal MHC Class II
```
mouse
  H2-IAb
  H2-IAd
  H2-IEd
```

Display binders ⦿ Show [1] best binder(s) per MHC

You can choose how many top binders to show per MHC, or use a binding percentile rank cutoff

○ Show below [20] percentile rank (1-100) per MHC

E-mail result ☐ Predictions are slow. For more than a few HLAs/MHCs, we recommend e-mailed result.

(Submit) (Reset)

ELF results 1:

Epitopes from our CTL database aligned to your query sequence

Bold **red** letters indicate residues that differ from the query sequence. The symbol ◄ means the HLA of the epitope matches one of your submitted HLAs. Click on the epitope to see full database entry. Click on "align" to align the epitope to the sequence database via QuickAlign.

Epitopes shown here are completely within the bounds of your query. Epitopes that overlap the ends of your query are included in the "View database records" links above.

[Download] this alignment in format [table ▲▼]

```
DTVLEDMNLPGRWKPKMIG
DTVLEEMNL A*6802    [align]  ◄
DTVLEDINL A*6802    [align]  ◄
DTVLEEWNL A*6802    [align]  ◄
DTVLEEMNL A68       [align]
DTVLEEMNL A28       [align]
DTVLEDMNL           [align]
    EEMNLPGRW B44        [align]
    EEINLPGKW B44        [align]
    EEMNLPGRW B*4402       [align]
    EEMNLPGRW B*4403       [align]
    EEMNLPGRW B18,B40,B44    [align]
    EDMNLPGRW           [align]
    EEMNLPGRW B*44       [align]
    EEINLPGKW B*4403      [align]
    EEMNLPGRW          [align]
        LPGRWKPKMI Cw3       [align]
        LPGRWKPKMI B7     [align]
```

Clicking on an epitope takes you to respective CTL or Helper epitope Database entries

Los Alamos
NATIONAL LABORATORY

ELF results 2:

Potential epitopes based on anchor residues

These peptides have C-terminal anchor residues, highlighted in **blue**, and internal anchors highlighted in **magenta**. These anchor residues match one or more motifs associated with the submitted HLA.

[Download] this alignment in format [table ▲▼]

```
DTVLEDMNLPGRWKPKMIG

DTVLEDMNL (A*0205 ........[L])

DTVLEDMNL (A*6802 .[TV]......[VL])

 TVLEDMNLP (A*0206 .[VQ].......)

   LEDMNLPGR (DRB5*0101,DRB5*0101 [FYLM]..[QVIM]....[RK])
```

Los Alamos
NATIONAL LABORATORY

ELF results 3:
Potential epitopes based on IEDB database MHC binding predictions, by Alexander Sette's group

Potential epitopes based on IEDB binding predictions

Top binders for each MHC are highlighted in **blue**.
Prediction method: IEDB recommended
Low percentile = good binders
Show up to 1 binder(s) per MHC

Class I

Selected allele(s): A*6802, B*1501

(Download) this alignment in format [table ⇕]

```
DTVLEDMNLPGRWKPKMIG (Click MHC to see full list of IEDB predictions for that MHC)
     DMNLPGRW        B*1501 (26)
      MNLPGRWK       A*6802 (3.0)
```

Clicking on MHC links to the full list of IEDB predictions for that MHC (see next slide)

Class II

Selected allele(s): DRB5*0101

(Download) this alignment in format [table ⇕]

```
DTVLEDMNLPGRWKPKMIG (Click MHC to see full list of IEDB predictions for that MHC)
   TVLEDMNLPGRWKPK     DRB5*0101 (17.17)
```

Los Alamos
NATIONAL LABORATORY

ELF results 3:
Potential epitopes based on IEDB database MHC binding predictions, by Alexander Sette's group

IEDB Analysis Resource

| Home | Help | Example | Reference | Download | Contact |

MHC-I binding predictions - Prediction Results

Input Sequences

#	Name	Sequence
1	sequence 1	DTVLEDMNLPGRWKPKMIG

Prediction method: IEDB recommended | Low percentile = good binders
Check to expanded the result: ☐

Allele	#	Start	End	Peptide Length	Sequence	Method used	Percentile Rank
HLA-B*15:01	1	6	13	8	DMNLPGRW	*NetMHCpan*	26
HLA-B*15:01	1	3	13	11	VLEDMNLPGRW	*NetMHCpan*	27
HLA-B*15:01	1	3	11	9	VLEDMNLPG	Consensus (*ANN,SMM,CombLib_Sidney2008*)	27.60
HLA-B*15:01	1	8	17	10	NLPGRWKPKM	*NetMHCpan*	31
HLA-B*15:01	1	7	17	11	MNLPGRWKPKM	*NetMHCpan*	35
HLA-B*15:01	1	2	9	8	TVLEDMNL	*NetMHCpan*	36
HLA-B*15:01	1	2	11	10	TVLEDMNLPG	*NetMHCpan*	47
HLA-B*15:01	1	4	11	8	LEDMNLPG	*NetMHCpan*	48

Alamos
IAL LABORATORY
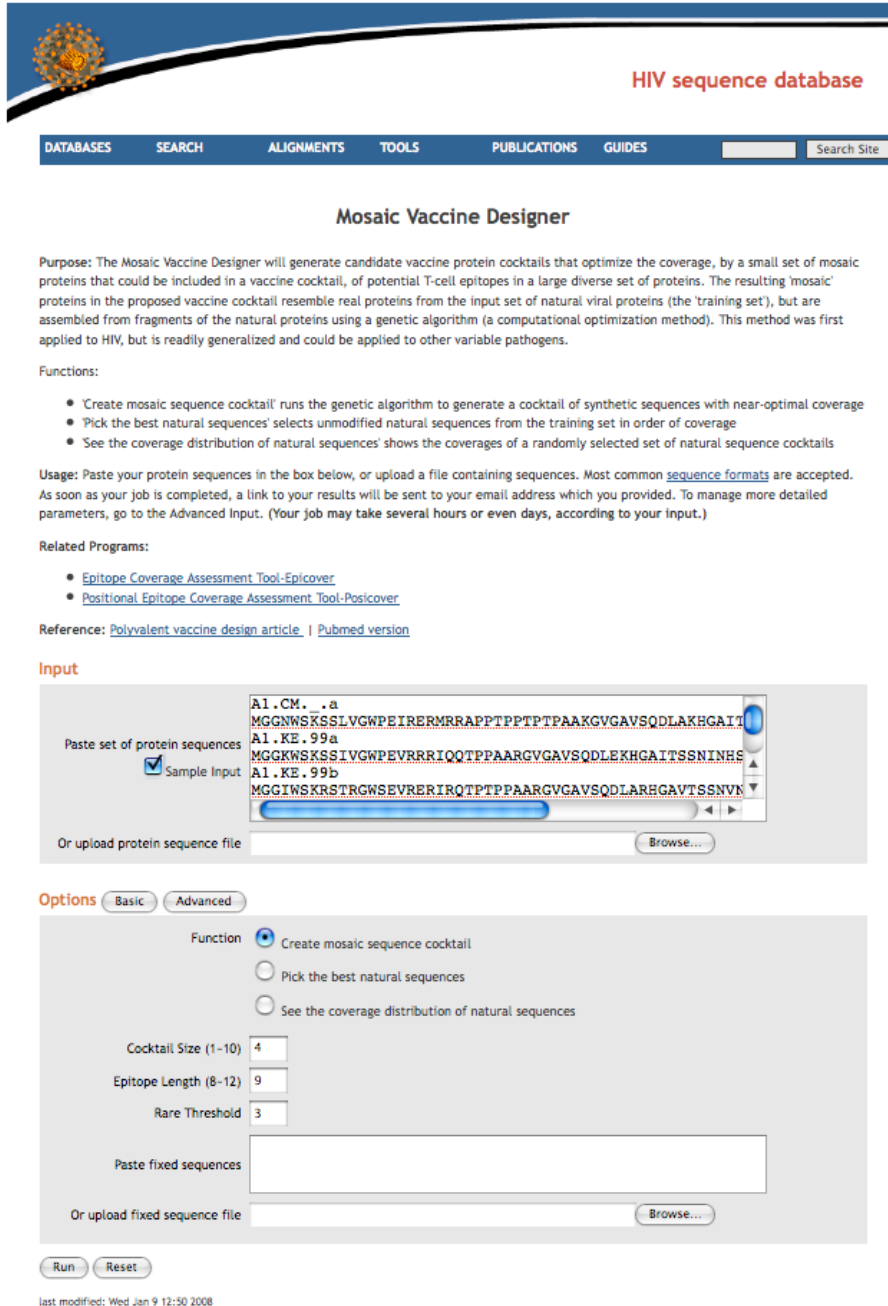
Mosaic vaccine tools

Mosaic Vaccine Designer: The Mosaic Vaccine Designer will generate candidate vaccine protein 'cocktails' that optimize coverage of potential T-cell epitopes found in a given background set of protein sequences.

Epitope Coverage Assessment: Alignment independent "n-mer" coverage of sequences by vaccines or peptides.

Positional Epitope Coverage Assessment: Alignment dependent coverage of sequences by vaccines or peptides.

Los Alamos
NATIONAL LABORATORY

Mosaic Vaccine Designer

Input: protein sequence set for a target population, does not need to be aligned.

Number of mosaic proteins in the set.

Epitope length.

Epitope Coverage Assessment - Epicover

Input

Use output from MakeVaccine tool

Provide a job number to access output from the Mosaic Vaccine Designer tool: [_____]

OR

Provide input sequences

Paste antigen protein sequence(s):
[Sample Input]

[   ]

upload more [+] antigen sequence files

and/or upload as files: [_____] (Browse...)

Paste test set protein sequences:

[   ]

upload more [+] test sequence files

and/or upload as files: [_____] (Browse...)

Options

Send results as an email instead of displaying in browser (useful in case of a browser time-out): ☐

Nominal epitope length: [9]

Maximum amino acid mismatches to score (range from 0): [2]

Minimum number of occurrences of a potential epitope in viral protein set to consider for coverage: [3]

Precision to use when reporting coverage: [4] decimal places

Advanced Options

Upload file of grouped sequence names [_____] (Browse...)

Report on subsets defined according to first [____] character(s) in sequence names

(Submit) (Reset)

Input:
Vaccine set
Test set

Can report on subsets defined according to the first several characters in sequence names or user-defined subsets

Los Alamos
NATIONAL LABORATORY

Epicover output

vaccine set	subset	subset count	Off-by-0	Off-by-1	Off-by-2	rare(<3,>1)	unique	absent
vaccine_set_from_user	**Total**	**63**	**0.6615**	**0.8914**	**0.9660**	**104**	**114**	**334**
vaccine_set_from_user	A	11	0.7232	0.9429	0.9935	47	36	334
vaccine_set_from_user	B	11	0.6378	0.8845	0.9755	25	19	334
vaccine_set_from_user	C	35	0.6921	0.8994	0.9637	51	45	334
vaccine_set_from_user	D	4	0.4217	0.7546	0.9443	4	9	334
vaccine_set_from_user	F	1	0.5300	0.7800	0.8600	4	5	334
vaccine_set_from_user	G	1	0.2597	0.7792	0.8312	0	0	334
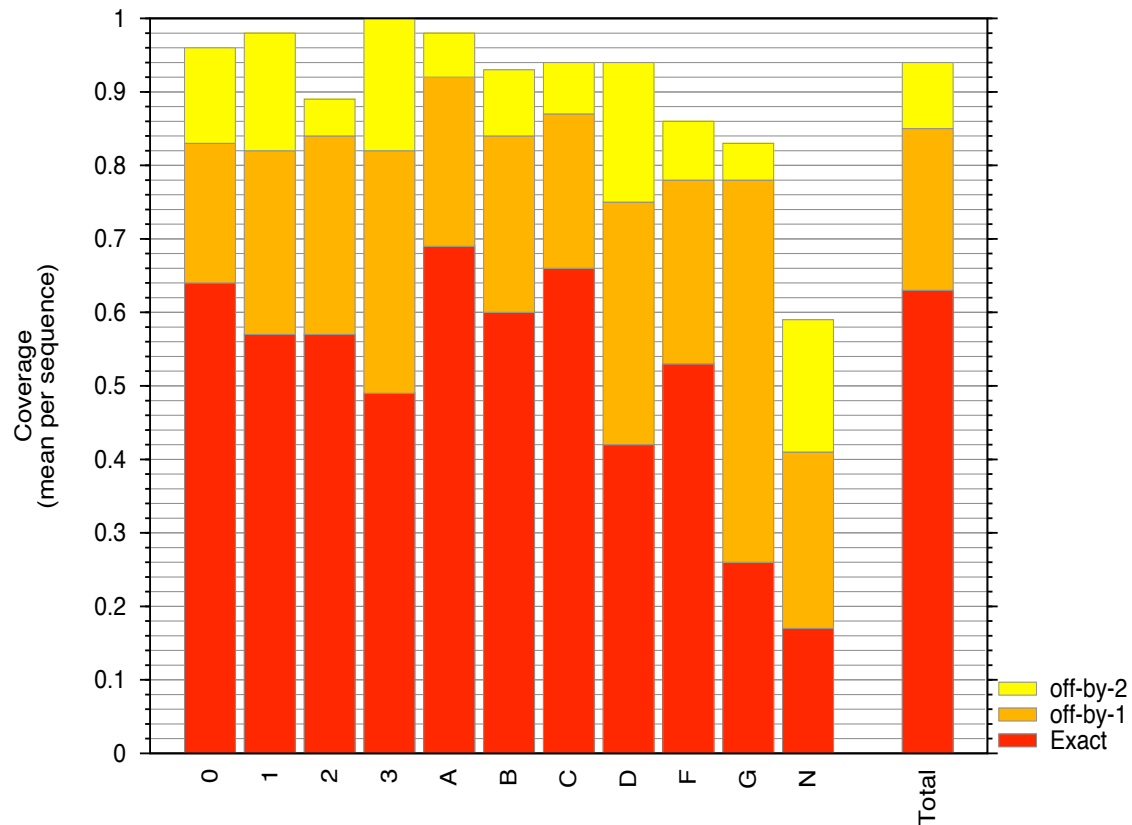
Positional Epitope Coverage Assessment - Posicover

Provide a job # from **Mosaic Vaccine Designer**:

(Only the antigen set is used. Provide the ALIGNED viral test set below)

AND/OR

Paste antigen protein set or peptide cocktail: (alignment not required) [Sample Input]

upload more [+] antigen files

and/or upload antigen file(s):

[Browse...]

Test set proteins

Paste **ALIGNED** test viral protein set: [Sample Input]
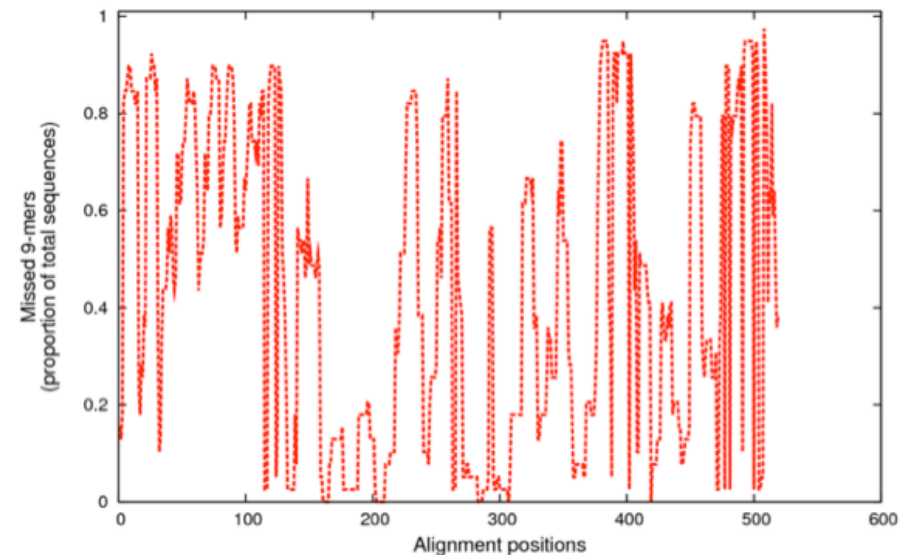
or upload an **ALIGNED** test proteins file:

[Browse...]

Input:

Vaccine set or Reagent set

ALIGNED test set

Los Alamos
NATIONAL LABORATORY

Examples of Posicover outputs

Matched 9-mers
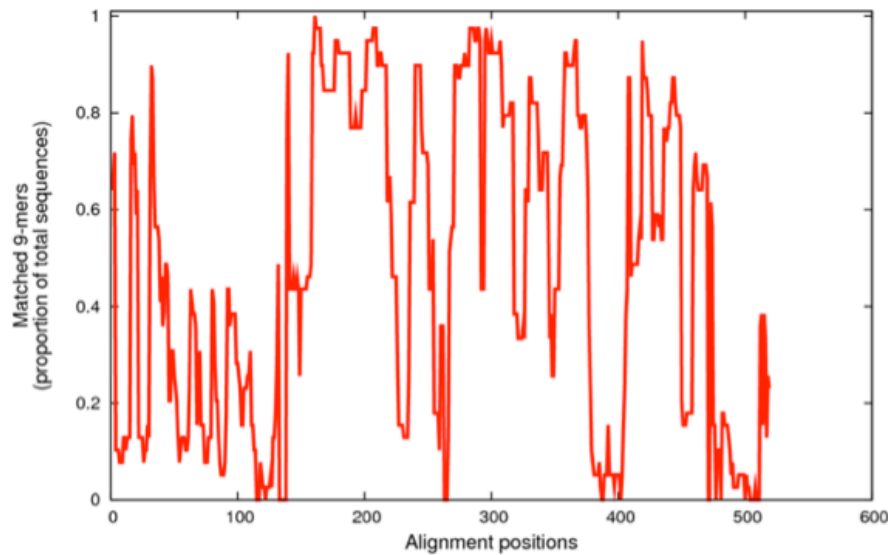
Missed 9-mers

Alignment positions
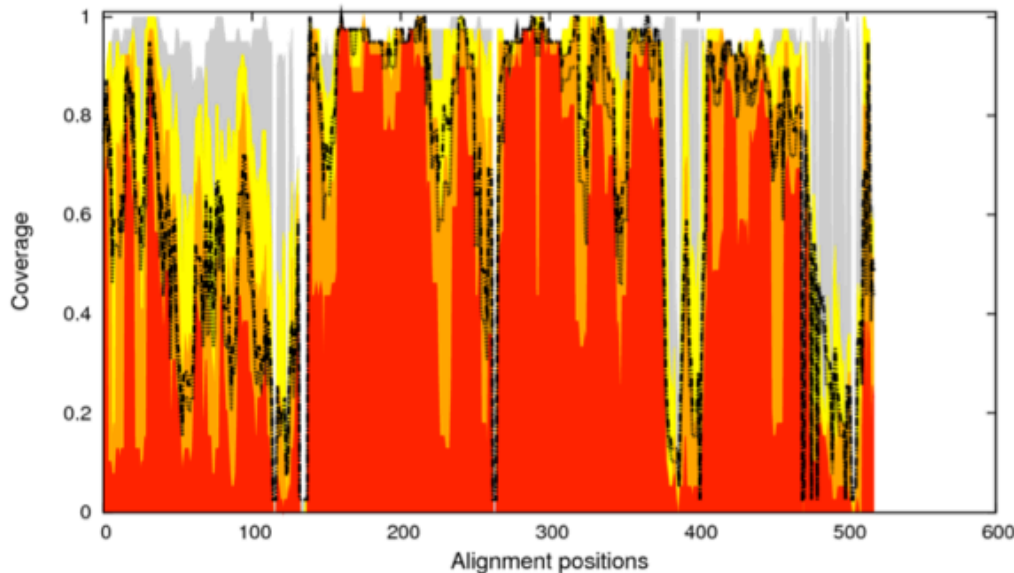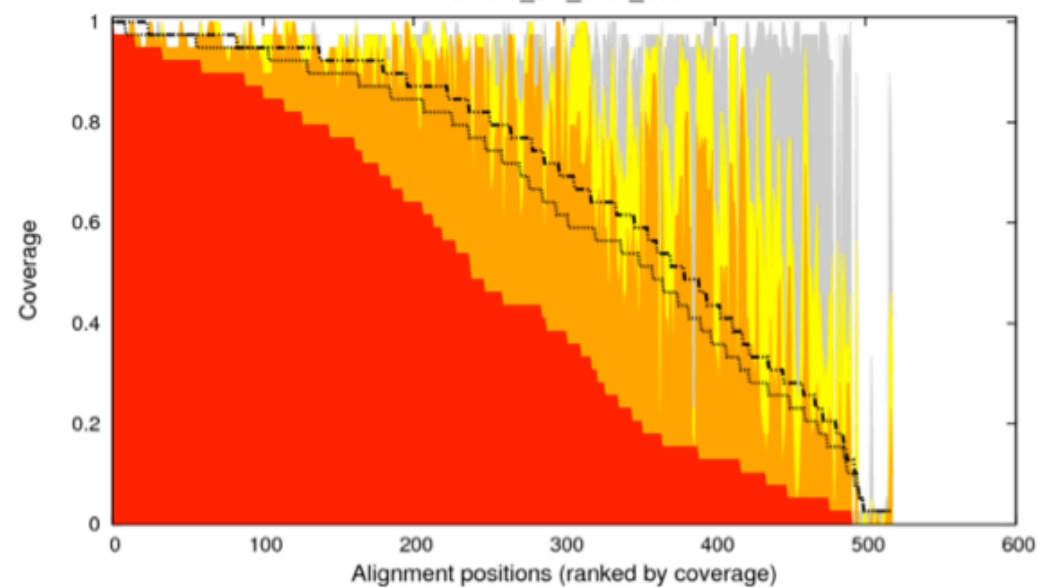
Examples of Posicover outputs

User's sequence alignment:

Each aa is represented as a single-colored square


Los Alamos
NATIONAL LABORATORY

Examples of Posicover outputs

Each amino acid is colored according to the
set of 9-mers that contain it:

Yellow: all 9-mers that overlap with
amino acid are perfectly matched in a test
vaccine set;

Increasingly red: fewer and
fewer matches in the overlapping set of 9-
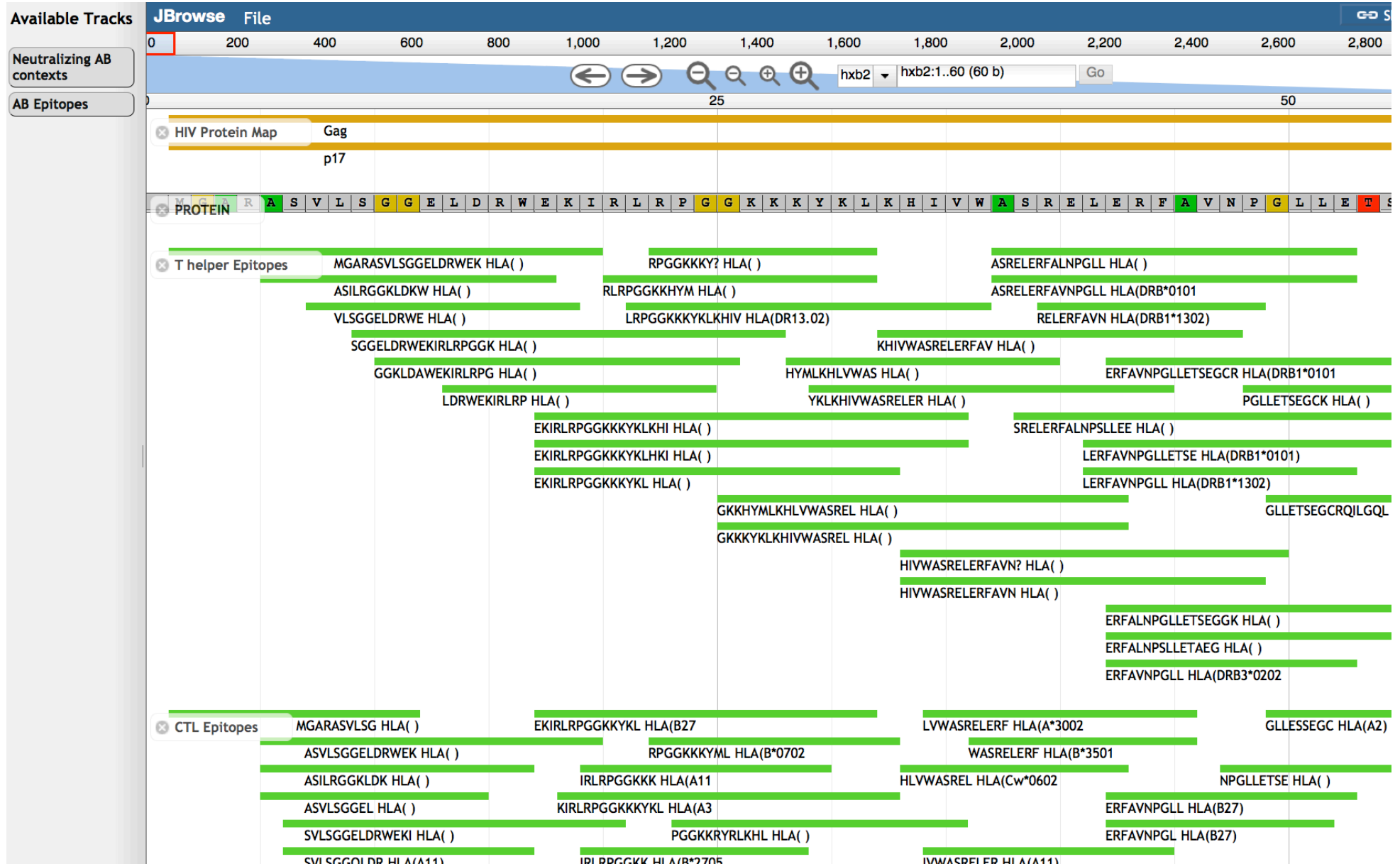mers that span the amino acid;

Black: amino-acid residues that are not
included in any vaccine set

0 1 2 3 4 5 6 7 8 9

unique 9-mer counts

Los Alamos
NATIONAL LABORATORY
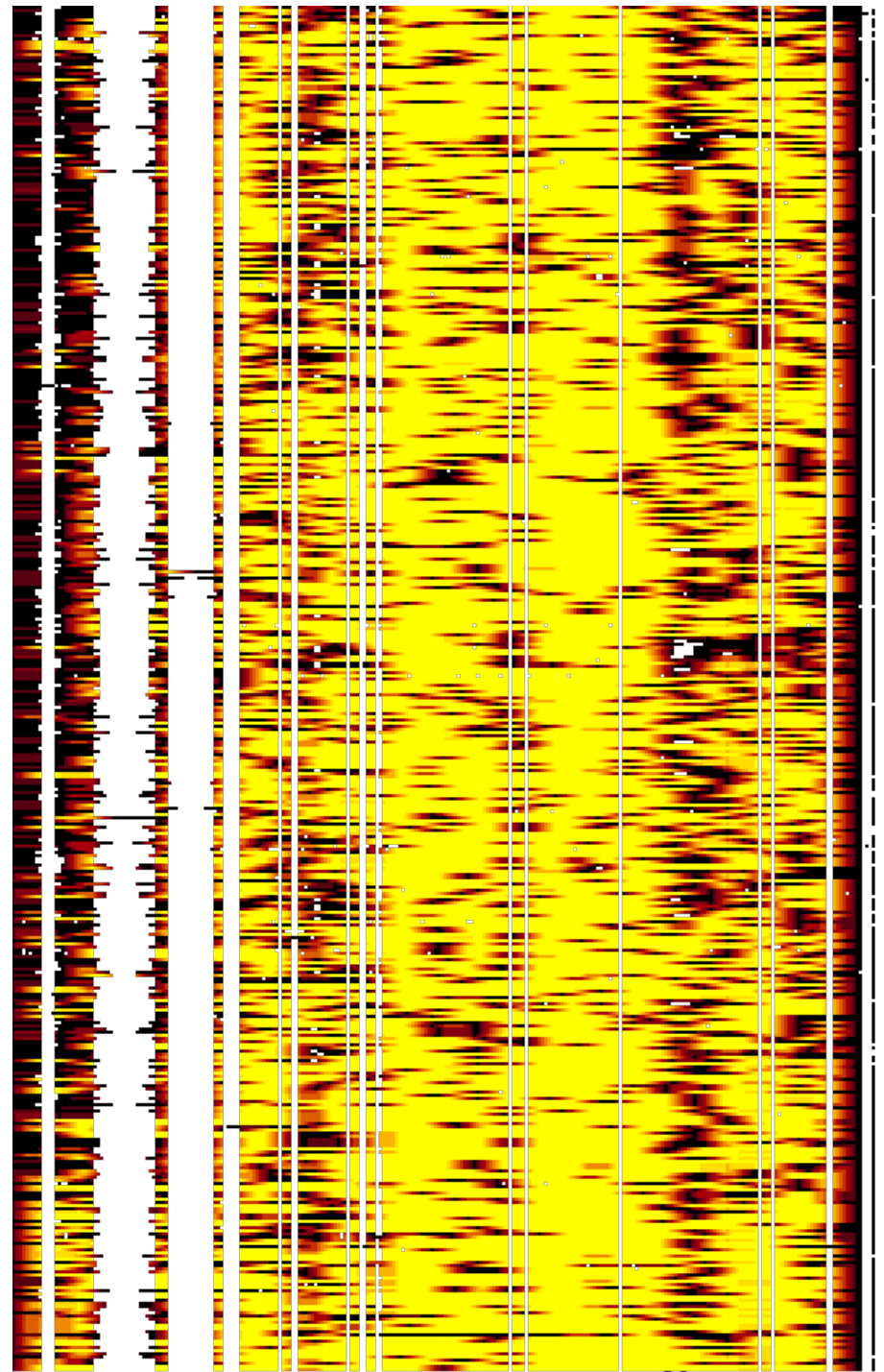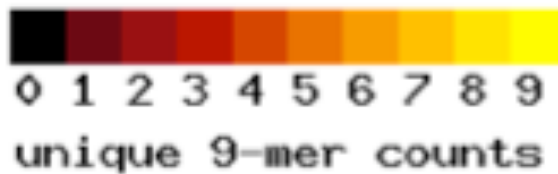
Coming soon:
HIV-1 genome browser

Coming soon:
HIV-1 genome browser

Los Alamos
NATIONAL LABORATORY

Please let us know if you have questions, comments or suggestions

seq-info@lanl.gov

Bette Korber: btk@lanl.gov
Will Fischer: wfischer@lanl.gov

Los Alamos
NATIONAL LABORATORY